Application of Linear Regression in Farmers' Income

Yichen Sheng*

Qingdao No. 9 Middle School, Qingdao, China

*Corresponding author: yichenelaineshengcqss@cseec.education

Abstract. This paper mainly discusses the application of linear regression model in the analysis of farmers' income and its influencing factors, and the application of Chinese tourism consumption in daily life. In the aspect of farmers' income, multiple linear regression model is widely used to analyze the influencing factors of farmers' income. Through the analysis of historical data and current situation, it is found that per capita wage income, per capita output value of agriculture, forestry, animal husbandry, fishery, per capita production cost expenditure, per capita transfer and property income have significant effects on farmers' income. These factors are considered to be the key to increase farmers' income, therefore, based on these analysis results, corresponding policy recommendations are put forward to promote the steady growth of farmers' income. In terms of tourism consumption, through the empirical analysis of multiple linear regression method, the research finds that rural tourism not only promotes the sales of agricultural and animal husbandry products, but also increases farmers' wage income and other sources of income, thus achieving a fundamental change in farmers' income. Linear regression model plays an important role in the analysis of farmers' income and its influencing factors, and provides a scientific basis for policy making. At the same time, the development of rural tourism has also been empirically supported by multiple linear regression method, which provides a new way for farmers to increase their income.

Keywords: Linear regression; Farmers' income; P-value test.

1. Introduction

It is known that the increase of farmers' income is directly related to the quality of life and consumption power of farmers. Only by continuously and rapidly increasing the income of rural residents, promoting consumption and investment in rural areas, and accelerating the narrowing of the gap between urban and rural areas can people promote economic growth and effectively reduce the proportion of low-income groups. Secondly, the increase of farmers' income will help narrow the gap between urban and rural areas and promote social equity and harmony. In addition, the increase in farmers' income can also enhance the self-development capacity of rural areas, enhance the level of agricultural modernization, and provide guarantees for national food security and sustainable development [1]. In the past few years, scholars' research on farmers' income has mainly focused on the application of econometric methods. For example, some studies have explored the main factors affecting farmers' income by establishing mathematical models and conducting multiple regression analysis, and quantitatively analyzed the extent of these factors. In addition, some studies use unconditional quantile regression method to discuss the new situation of rural residents' income distribution and the reasons for its change. These studies not only reveal the factors affecting farmers' income, but also provide a scientific basis for policy making [2].

Regarding farmers' income, it can also be investigated and analyzed by the following methods. The first, Rural Household Survey: The National Bureau of Statistics calculates rural income through the "Rural Household Survey", which reflects the income and quality of life of rural residents through a comprehensive survey of the production, consumption, income, accumulation, and social activities of rural households. Secondly, econometric methods: Based on the econometric model, the influencing factors of farmers' income and their mechanisms are studied. For example, multiple linear regression is used to analyze the main influencing factors of farmers' per capita net income. Thirdly, artificial neural network model: The artificial neural network method is used to forecast farmers' income, and there are many application cases in the field of forecasting [3]. Forth, theil index method: used to measure the change of farmers' income gap and analyze its causes. Finally, confirmatory

factor analysis: Explore the grade evaluation, regional differences, and spatio-temporal evolution of farmers' income quality.

Linear regression analysis can be divided into two types: univariate regression and multiple regression. Univariate regression analysis involves only one independent variable and one dependent variable, and the relationship between them is studied by establishing linear equations. Multiple regression analysis involves multiple independent variables and one dependent variable, and studies the relationship between multiple independent variables and dependent variables by establishing linear equations [4]. The basic idea of multiple regression analysis is to predict or estimate the value of the dependent variable through the optimal combination of several independent variables. This method can explain changes in dependent variables more fully because it considers the combined influence of multiple factors. The goal of multiple regression analysis is to identify a set of regression coefficients that represent the extent to which each independent variable influences the dependent variable, considering the effects of other independent variables. In practical application, multiple regression analysis needs to meet some preconditions, such as the absence of multicollinearity between independent variables. Multicollinearity means that there is a high correlation between multiple independent variables, which affects the stability and explanatory power of the regression model. Therefore, dealing with multicollinearity is an important step in multiple regression analysis. Multiple regression analysis is widely used in various fields, such as economics, social sciences, and biostatistics. Through multiple regression analysis, researchers can more accurately predict and explain complex phenomena, to provide scientific basis for decision-making [5].

The aim of this paper is organized as follows. Section 2 will give linear regression model and data of farmers' income and its influencing conditions. Section 3 will propose the OLS data analysis, P-value detection and linear regression are used in the study of domestic tourism consumption. Finally, Section 4 is devoted to the conclusion of this paper.

2. Methodology

2.1. Linear Regression

Linear regression is a regression analysis method commonly used in statistics and machine learning. Its basic principle is to predict and explain by establishing the linear relationship between the independent variable and the dependent variable. Specifically, linear regression attempts to find the best linear function that most accurately predicts the value of the dependent variable given the value of the independent variable. The advantages of linear regression lie in its simple model form, high computational efficiency, and good interpretability of the results. The degree of influence of each feature on the target variable can be intuitively expressed by the regression coefficient. However, linear regression also has its limitations, such as it assumes a linear relationship between independent and dependent variables, which may not be applicable in some cases. In conclusion, linear regression is used to predict and interpret by fitting a linear relationship between the independent variable and the dependent variable, and the least squares method or other optimization algorithms are used to find the optimal model parameters, to minimize the difference between the predicted value and the actual value [6].

The basic assumption of linear regression is that there is a linear combination of one or more independent variables (x) that can well predict the value of the dependent variable (y). For example, for unary linear regression, the model can be expressed as

$$y = w \cdot x + b \tag{1}$$

where w is the weight and b is the bias term. For multiple linear regression, it is necessary to consider the linear combination of several independent variables. To achieve this, linear regression typically uses least squares to find the optimal solution, minimizing the sum of squares of the errors between the dependent and independent variables. In addition, optimization algorithms such as

gradient descent method and Newton method can be used to solve the optimal weight vector and bias term.

2.2. Data Resource

There are some variables about the farmers' incomes and the conditions that affect them. The first is farmers' per capita net income y (explained variable). The second is per capita wage income x_1, per capita output value of agriculture, forestry, animal husbandry and fishery x_2, per capita production cost expenditure x_3, per capita transfer and property income x_4 (explanatory variables). Set up an econometric model to explain the relationship between farmers' per capita net income and per capita wage income, per capita transfer of production expenses and per capita property income. There are four variables in the creation process of this mode. They are China's per capita wage income, per capita production expenditure of agricultural, forestry, animal husbandry and fishery output, per capita transfer, and property income from 1995 to 2009. Therefore, for time series data, the last data in 2009 is used as forecast comparison data, and the collected data is shown in Table 1 [7].
If the farmers' per capita net income is correlated with five explanatory variables, the sample regression model is

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + e_i. \quad (2)$$

Table 1. The data collection of the variables X1, X2, X3, and X4 stands for er capita wage income, per capita output value of agriculture, forestry, animal husbandry and fishery, per capita production cost expenditure, per capita transfer and property income.

obc	Y	X1	X2	X3	X4
1995	1,302,950	2,342,100	1,977,920	4,654,000	9,361,000
1996	1,607,720	3,463,500	2,263,550	4,568,800	1,072,100
1997	1,808,750	4,054,900	2,461,090	4,801,500	7,797,000
1998	1,863,060	4,571,700	2,463,460	4,401,500	1,046,100
1999	1,900,290	4,706,900	2,468,030	4,091,000	1,200,800
2000	1,934,570	5,478,300	2,416,650	4,491,100	1,274,100
2001	2,020,000	6,106,500	2,482,250	4,860,800	1,417,900
2002	2,118,000	7,076,800	2,570,050	5,265,100	1,661,700
2003	2,117,000	8,189,200	2,563,820	5,176,800	1,408,700
2004	2,449,300	8,846,200	3,212,360	6,989,000	1,653,400
2005	2,641,000	1,010,050	3,236,390	8,974,100	1,877,100
2006	2,969,100	1,184,110	3,449,630	1,035,960	2,266,700
2007	3,556,300	1,470,050	3,974,380	1,119,540	3,278,200
2008	4,202,500	1,737,840	4,666,690	1,395,220	4,373,600
2009	4,504,300	1,882,420	4,846,540	1,334,260	4,774,500

3. Result

3.1. Through OLS Analysis of Farmers' Income Results and P-value Test

This result was obtained by using the OLS method through 15 sets of data. Ordinary least squares (OLS) regression is an optimization strategy that aims to find a straight line in a linear regression model that is as close as possible to the data point. According to the coefficient and the sample regression model, the results of farmers' per capita net income are [8]

$$Y_i = 33.632 + 0.659X_1 + 0.59X_2 - 0.274X_3 + 0.152X_4 \quad (3)$$

and

$$SE = (186.261)\ (0.1815\ (0.1245)\ (0.2037)\ (0.5699) \quad (4)$$

The R-square is close to one. Therefore, the fitting degree of the model is closer to the reality. The summary of the statistics is shown in Table 2. The Linear Regression of China's per capita wage income is shown in Fig. 1.

Table 2. Summary Statistics.

	Y	x_1	x_2	x_3	x_4
Count	1.50000e+01	1.500000e+01	1.500000e+01	1.500000e+01	1.500000e+01
Mean	2.46666e+06	4.141371e+06	3.003521e+06	4.210579e+06	2.964193e+06
Std	9.476277e+05	2.585149e+06	8.812211e+05	2.228307e+06	2.573991e+06
Min	1.302950e+06	1.010050e+06	1.977920e+06	1.035960e+06	1.046100e+06
25%	1.881675e+06	1.810130e+06	2.462275e+06	2.743110e+06	1.341400e+06
50%	2.117000e+06	4.054900e+06	2.563820e+06	4.568800e+06	1.661700e+06
75%	2.805050e+06	5.792400e+06	3.343010e+06	5.018800e+06	3.825900e+06
Max	4.504300e+06	8.846200e+06	4.846540e+06	8.974100e+06	9.361000e+06

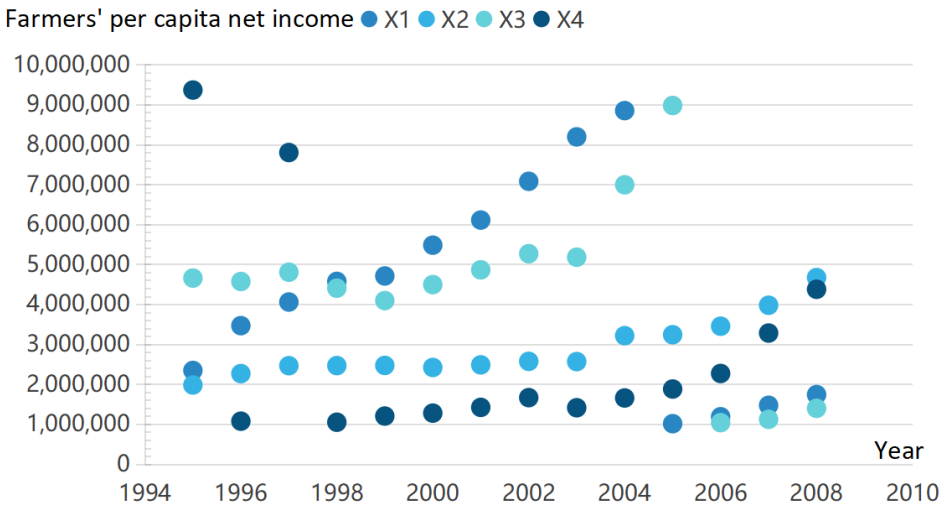

Fig. 1 The Linear Regression of China's per capita wage income, per capita production expenditure of agricultural, forestry, animal husbandry and fishery output, per capita transfer and property income from 1995 to 2009.

When the P-value is greater than 0.1, there is no linear relationship between them. From the results, its clearly to see that the values of x_1 and x_4 are much greater than 0.1, so they are useless. Among them, the most useful is the x_2 value, and the second one is constant. Next is the diagram about Residuals. Except for the nadir, all the other points are pretty much symmetric about 0. The OLS Regression Results is shown in Table 3.

Table 3. The OLS Regression Results.

	coef	Std_err	t	P>|t|	[0.025	0.975]
const	-5.145e+05	1.83e+05	-2.816	0.018	-9.22e+05	-1.07e+05
x_1	0.0028	0.013	0.216	0.834	-0.026	0.032
x_2	1.0382	0.038	27.313	0.000	0.954	1.123
x_3	-0.0270	0.015	-1.821	0.099	-0.060	0.006
x_4	-0.0120	0.011	-1.059	0.314	-0.037	0.013

The government should take various measures to ensure the continuous increase of farmers' income. First, the process of labor transfer should be continued, while measures should be taken to

compensate for its possible negative impact on agricultural production. Second, people will increase policy support to strengthen farmers, benefit them and enrich them, so that farmers' wage income, business income, property income and transfer income can continue to grow. In addition, people will improve the social security system and deepen the reform of the property rights system to increase the transfer and property income of urban and rural residents. Finally, people should promote the growth of farmers' wage income, transfer income and property income through fiscal agricultural expenditure.

3.2. Study on the Influence of Linear Regression on Domestic Tourism Consumption

As a major tourism country in the world, China has many factors affecting tourism consumption, including political aspects, economic aspects, social aspects and environmental aspects. By combing domestic and foreign literatures and removing factors and indicators that cannot be quantitatively analyzed, this paper selects representative factors including the average disposable income of urban and rural residents and the number of tourists to conduct regression research on domestic tourism income, and expresses the influence of each factor on tourism consumption in the form of correlation functions. At the same time, since the total amount of consumption expenditure and consumption income should be equal on a macro level, the dependent variable of this paper selects domestic tourism income y_i to represent tourism consumption and x_i to represent a specific influencing factor [9].

Of note is that there are some special variables, Since per capita disposable income can be divided into urban disposable income and rural disposable income, its variables are set as x_1 and x_2 respectively. x_3 represents the impact of the number of tourists on the domestic tourism income. The China's domestic tourism related statistics and related statistical variables are shown in Table 4.

Table 4. China's domestic tourism related statistics and related statistical variables from 2007 to 2019.

Year	Domestic tourism revenue (100 million yuan)	Per capita Disposable income of urban residents (Yuan)	Per capita disposable income of rural residents (Yuan)	Number of domestic tourists (100 million)
2007	249,530.60	13,786	4327	16.1
2008	300,670	15,781	4999	17.1
2009	340,507	17,175	5435	19
2010	397,980	19,109	6272	21
2011	471,564	21,810	7294	26.4
2012	519,322	24,565	8389	29.6
2013	568,845	26,955	9430	36.11
2014	636,463	28,844	10,489	40
2015	689,053	31,195	11,422	41
2016	743,586	33,616	12,363	44.44
2017	832,036	36,396	13,432	50.01
2018	9,192,816	39,251	14,617	55.39
2019	99,022,656	42,359	16,021	60.06

Table 5 shows the results of the model ANOVA. According to Table 5, the F-value of the model is 863.656 and the P-value is 0.000. When the significance level is 0.05, the author rejects the null hypothesis because the P-value is less than 0.05, which indicates that at least one of the three variables - GDP, number of tourists and average household income - has a significant impact on domestic tourism income.

Table 5. Analysis of variance

Model		Sum of squares	df	Mean square	F	statistical significance
1	Regression	256880005233.044	4	64220001308.261	863.656	0.000^{b}
	Residual error	371791394.976	5	74358278.995		
	Total	257251796628.020	9			

In the regression coefficient shown in Table 6, after logarithm of x_1, x_2, x_3 is taken, all of them pass the T test. From this, it can be concluded that the multiple linear regression model affecting domestic tourism can be expressed as:

$$Y = -63.1513 + 35.1168x_1 - 32.6556x_2 + 4.9490x_3 \quad (5)$$

Table 6. Regression coefficient table

	Coefficient	Standard error	T-test
Constant term	63 1513	39.9897	-1.58
Inurgdi	35.1168*	16.1986	2.17
Inesgdi	32.6556*	16.7953	-1.94
Inpeo	4.9490**	1.3485	3.67

According to the multiple linear regression analysis, it can be seen that the variable urban per capita resident income has a greater impact on the domestic tourism income, while the variable rural per capita income has a smaller impact on the domestic tourism number than the variable and the number of tourists. With the continuous development of domestic economy, the increase of per capita income and other factors, one can see that domestic tourism income will continue to increase. In view of the above analysis and the corresponding development facts, some relevant suggestions are provided to promote the further development of domestic tourism. First, it can be seen from the analysis of this paper that the income of urban residents has a significant impact on domestic tourism income, so it is necessary to strengthen the publicity of urban residents' tourism and appropriately develop scenic spots close to their residents. Although the per capita income level of rural residents has a relatively small impact on tourism income, the difference in domestic income distribution should be narrowed, such as vigorously developing rural happiness, promoting suburban short trips, and so on, so as to strengthen the driving role of rural residents on tourism economy.

Second, it is necessary to continuously promote the improvement of economic aggregate, strengthen the overall development of urban and rural areas, and improve the per capita disposable income level, so that people will be rich from the material level to the spiritual level of pursuit, and increase the investment in tourism consumption. The government should also encourage the development of rural and urban-rural tourism, and give more policy support and help to such development. Third, promote the common development of different industries. All along, the related development of tourism is not independent, and the development of industry can drive the development of agriculture. From the above regression analysis, it can be seen that there is still a large blank area in rural tourism, and the rural market has great potential for development. Therefore, the development of agriculture will inevitably greatly promote the development of tourism and increase the income of domestic tourism.

4. Conclusion

This paper introduces what is linear regression, how to use linear regression to solve the problems in people's daily life such as per capita wage income, per capita agricultural and forestry production expenditure, animal husbandry, fishery output, per capita transfer from 1995 to 2009, property income, OLS, and how to use P-value test to verify the calculation results. At the same time, the influence of

linear regression on domestic tourism consumption is also studied. It proves the application of linear regression in life. In the future, linear regression will be applied to more fields, which is conducive to predicting stock prices, currency exchange rates, housing prices, etc. By collecting historical data, establishing models and conducting evaluation and optimization, future development trends can be predicted. It can also predict the risk factors and the development trend of the disease. Through statistical analysis of economic data, people can study the linear relationship between different variables, so as to predict the future development of the economy.

References

[1] Suzuki Sei. Half Landau–Zener Ramp to a Quantum Phase Transition in a Dissipative Single Spin Model. The European Physical Journal B, 2024, 97(8): 122.

[2] Chen Xingxing. Research on influencing factors of domestic tourism consumption based on multiple linear regression analysis. Advances in Social Science, 2020, 9(9): 1530-1536.

[3] Lili Ilma, and Anxhela Kosta. Applying an Ordinary Least Squares (OLS) Regression Model on Processed Air Quality and Environment Data. British Journal of Environmental Sciences, 2024, 12(2): 49–58.

[4] Gou Jianwei, Qian Yaofei, and Wang Zeqing. The method and application of quantile regression model in high-dimensional financial data analysis. China Academic Journal Electronic Publishing House, 2019, (07): 39-41

[5] Li Wenyue, Li Jungang. Based on linear regression model, this paper makes an empirical analysis of China's GDP growth. Statistics and Application, 2022, 12(29): 1499.

[6] Su Qing. Prediction and Application of Interval Value of Security Price Based on Recurrent Neural Network and Regression Analysis. Zhengzhou University, 2021.

[7] Wang Lijun. The Para Meter Estimates of Fuzzy Linear Regression Model and the Application in finance. Dong Hua University, 2006.

[8] Chen Yutian. Comparison of linear and nonlinear multi-factor stock selection models in a-share market. Harbin Institute of Technology, 2019.

[9] Zhou Yi. Regression Analysis and Driving Force Model Building of CO2 Emissions in China. Scientific Reports, 2021, 11(1): 6715.